1. DISTRIBUCIONES DE PROBABILIDAD DE VARIABLES DISCRETAS.

Definición

La distribución de probabilidad de una variable aleatoria discreta es una tabla, gráfica, fórmula o cualquier otro medio que se use para especificar todos los valores posibles de una variable aleatoria discreta junto con sus probabilidades respectivas.

Ejemplo

Supongamos que una enfermera tiene a su cargo 50 familias. Construyamos la distribución de probabilidad X, el número de niños por familia para esa población.

Distribución de probabilidad del número de niños por familia en una población de 50 familias.

x	Frecuencia de la ocurrencia de x	$P(X = x)$
0	1	1/50
1	4	4/50
2	6	6/50
3	4	4/50
4	9	9/50
5	10	10/50
6	7	7/50
7	4	4/50
8	2	2/50
9	2	2/50
10	1	1/50
	50	50/50

Tabla 1.1

En la tabla anteriormente expuesta se observan una serie de características comunes a toda tabla de distribución de probabilidad:

- Los valores P son todos positivos
- Todos ellos son menores que 1
- Su suma es igual a 1.

Además podemos observar cómo la probabilidad de un determinado suceso coincide con la frecuencia relativa de dicho suceso. Así podemos contestar fácilmente a cualquier pregunta que nos hagan acerca de dicha población. Por ejemplo:

- ¿ Cúal es la probabilidad de que la familia tenga 4 niños?

Solución: 9/50

- ¿ Cúal es la probabilidad de que la familia elegida al azar tenga 4 ó 5 niños?

Solución:

Debemos aplicar la regla de la adición. Dado que la probabilidad de tener 4 hijos es 9/50 y la de tener 5 es 10/50 podemos concluir que la probabilidad de tener 4 ó 5 hijos es 19/50.

Otra de las formas de representar las distribuciones de probabilidad de variables discretas es, según vimos en la definición, mediante gráficos como el siguiente del cual podemos obtener la misma información que la obtenida mediante la tabla 1.1

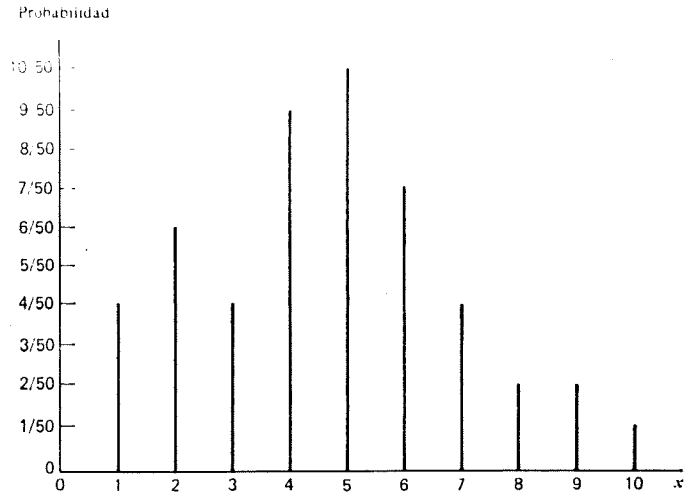


Figura 1.1 Representación gráfica de una distribución de probabilidad del número de niños por familia para una población de 50 familias.

Distribuciones acumuladas

A veces resulta más conveniente trabajar con la distribución de probabilidad acumulada de una variable aleatoria. La distribución de probabilidad acumulada de una variable aleatoria discreta cuya distribución de probabilidad vimos en la tabla 1.1 puede obtenerse, sumando sucesivamente las probabilidades, P(X=x), que se dan en la última columna. Así obtenemos la siguiente distribución:

x	Frecuencia de la ocurrencia de x	$P(X = x)$	$P(X \leq x)$
0	1	1/50	1/50
1	4	4/50	5/50
2	6	6/50	11/50
3	4	4/50	15/50
4	9	9/50	24/50
5	10	10/50	34/50
6	7	7/50	41/50
7	4	4/50	45/50
8	2	2/50	47/50
9	2	2/50	49/50
10	1	1/50	50/50
	50	50/50	

Tabla 1.2. Distribución de probabilidad acumulada del número de niños por familia en una población de 50 familias.

Con la anterior distribución podemos contestar fácilmente a preguntas como éstas:

- ¿ Cúal es la probabilidad de que una familia elegida aleatoriamente tenga menos de 6 niños ? 34/50
- ¿ Cúal es la probabilidad de una familia elegida al azar de entre las 50 tenga 5 ó más niños ? 1-(24/50)
- ¿ Cúal es la probabilidad de que una familia elegida al azar tenga entre 3 y 6 niños inclusive ? (41/50-11/50)

La distribución de probabilidad dada en la tabla 1.1 se desarrolló a partir de la experiencia real. Sin embargo, las distribuciones de probabilidad de muchas variables de interés pueden determinarse o suponerse sobre la base de consideraciones teóricas. Dos de estas distribuciones teóricas son la binomial y la de Poisson.

2. LA DISTRIBUCION BINOMIAL.

Introducción

La distribución binomial es una de las distribuciones de probabilidad que se encuentran con más frecuencia en la estadística aplicada. La distribución se obtiene de un proceso conocido como ensayo de Bernoulli, en honor del matemático suizo James Bernoulli (1654-1705) , quien realizó importantes contribuciones en el campo de la probabilidad incluyendo, en particular, a la distribución binomial. Cuando un sólo ensayo de algún proceso o experimento puede conducir sólo a uno de dos resultados mútuamente exclusivos, tales cono enfermo o saludable, masculino o femenino, el ensayo se conoce como ensayo de Bernoulli.

Condiciones del proceso de Bernoulli

a) Cada ensayo conduce a uno de dos resultados posibles, mútuamente exclusivos. Uno de los dos resultados posibles se denota (arbitrariamente) como éxito y el otro se denota como fracaso.

b) La probabilidad de éxito, denotada por p, permanece constante de ensayo a ensayo. La probabilidad de fracaso, 1-p, se denota por q.

c) Los ensayos son independientes, es decir, el resultado de cualquier ensayo particular no es afectado por el resultado de cualquier otro ensayo.

Parámetros de la distribución

La distribución binomial tiene 2 parámetros. Son parámetros en el sentido de que bastan para especificar a una distribución binomial. La media y la varianza de la distribución binomial son $\mu = np$ y $\sigma = np(1-p)$. Si denotamos el tamaño de la población como N, se conviene que para utilizar el modelo binomial n debe ser pequeño comparado con N. Algunos autores dicen que n es pequeño respecto a N, si N es por lo menos 10 veces mayor que n.

Un ejemplo práctico

Supóngase que en cierta población, el 52% de todos los nacimientos registrados son varones. Esto se interpreta en el sentido de que la probabilidad de un nacimiento registrado de varón es 0.52. Si de esta población se seleccionan al azar 5 registros de nacimientos, ¿ cúal es la probabilidad de que exactamente 3 de los registros sean de nacimientos de varones ?.

Vamos a considerar como éxito el nacimiento de varón que designaremos como 1 y el fracaso vamos a designarlo como 0. Supongamos que la sucesión de sexos fuese MFMMF, es decir, 10110. La probabilidad de MFMMF será pqppq,es decir, q^2p^3. Además de la sucesión anteriormente expuesta pueden ocurrir las sucesiones siguientes:

Número	Sucesión
2	11100
3	10011
4	11010
5	11001
6	10101
7	01110
8	00111
9	01011
10	01101

Cada una de las sucesiones anteriormente expuestas tienen una probabilidad de q^2p^3, la probabilidad calculada para la

primera sucesión. La cuestión ahora se convierte en la de ¿ cúal es la probabilidad de obtener la sucesión 1 ó la 2 ó ... la número 10 ? Por la regla de la adición será $10q^2p^3$.

Ahora puede darse respuesta a la pregunta original: ¿ cúal es la probabilidad en una muestra de tamaño 5, extraída de la población especificada, de observar 3 éxitos y 2 fracasos ?. La respuesta es 0.32.

Fácilmente puede anticiparse que, a medida que aumenta el tamaño de la muestra la enumeración del número de sucesiones se hace cada vez más difícil y tediosa. Lo que se necesita es un método fácil de contar el número de sucesiones. Por la ecuación n!/x! (n-x)! el número de sucesiones en el presente ejemplo es 5!/2! 3! = 120/12 = 10. El resultado de la pregunta original puede darse de forma global si aplicamos uno de los términos del binomio de Newton. Aplicaríamos la fórmula:

$$f(x) = \binom{n}{x} q^{n-x} p^x$$

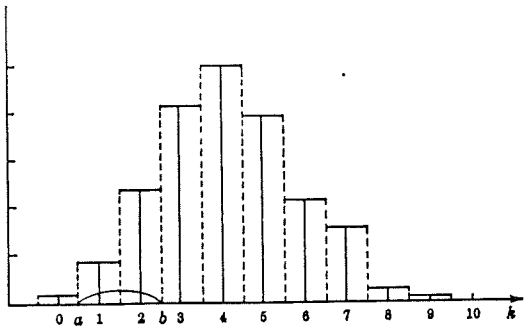
La distribución normal como una aproximación de la distribución binomial

Como hemos visto en la explicación anterior, el cálculo de las probabilidades binomiales es tarea difícil en muchos casos. Afortunadamente, en ciertas condiciones, la distribución binomial se aproxima a la distribución normal que es más fácil de calcular. Vamos a explicar la relación entre las distribuciones normal y binomial y demostraremos que la distribución binomial se parece a la distribución normal cuando n es muy grande y ni p ni q son próximos a 0. En estas condiciones, la variable sigue aproximadamente una distribución normal de $\mu = 0$ y $\sigma = 1$. Hay que tener en cuenta que para utilizar correctamente esta transformación de una variable discreta (con distribución binomial) en una variable continua Z (con distribución normal) es necesario hacer una corrección de continuidad. Sin embargo, es aceptable esta transformación cuando p es menor o igual que 0.5 y n x p mayor que 5.

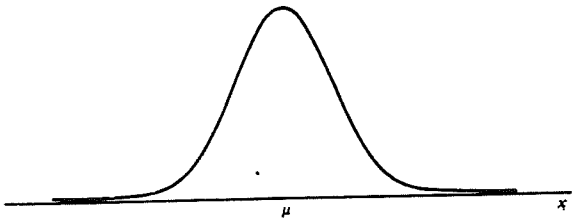
Supóngase que en una población de estudiantes $\pi = 0.40$ fuman. Se toma una muestra de tamaño 10 ensayos de Bernoulli repetidos. Sea k el número de fumadores (éxitos) de la muestra. Las probabilidades binomiales se muestran en la tabla siguiente:

k	$p = k/n$	$b(k; n = 10, \pi = 0.4)$
0	0	0.006
1	0.1	0.040
2	0.2	0.121
3	0.3	0.215
4	0.4	0.251
5	0.5	0.201
6	0.6	0.111
7	0.7	0.042
8	0.8	0.011
9	0.9	0.002
10	1.0	0.000
		1.000

En la siguiente figura se muestra un histograma de esas probabilidades.



Podemos ver que a medida que n aumenta, la amplitud de los rectángulos y los escalones del histograma disminuyen. La línea quebrada se acerca a una línea suave. Establecemoas, sin probarlo rigurosamente, que cuando n--∞ esta curva suave se aproximará a una curva normal. Es decir, la distribución binomial se aproximará a una distribución normal.



3. LA DISTRIBUCION DE POISSON.

Introducción

Esta distribución, usada ampliamente en medicina y biologia, se llama asi en honor del matemático francés Simeon Denis Poisson (1781-1840). La distribución de Poisson puede considerarse como una forma límite de la distribución binomial cuando n es grande y p muy pequeño. Sin embargo, también se puede considerar en si misma observando un proceso de Poisson.

El proceso de Poisson

a) Las ocurrencias de los eventos son independientes. La ocurrencia de un evento en un intervalo de espacio o tiempo no tiene efecto sobre la probabilidad de una segunda ocurrencia del evento en el mismo, o cualquier otro intervalo.

b) Teóricamente, debe ser posible un número infinito de ocurrencias del evento en el intervalo.

c) La probabilidad de la ocurrencia única del evento en el intervalo dado es proporcional a la longitud del intervalo.

d) En cualquier porción pequeña del intervalo, se desprecia la probabilidad de más de una ocurrencia del evento.

Una particularidad interesante de la distribución de Poisson es que la media y la varianza son iguales.

La distribución

Si x es el número de ocurrencias de algún evento aleatorio en un intervalo de tiempo o espacio (o algún volumen de materia) la probabilidad de que ocurra x está dada por:

$$f(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

La letra griega lambda se llama parámetro de la distribución y es el número promedio de ocurrencias del evento aleatorio en el intervalo. El símbolo e es la constante 2.7183.

Un ejemplo práctico

Poisson fue un médico militar francés que estudió la probabilidad de que un soldado muriera en el campo de batalla a consecuencia de los golpes propiciados por un caballo. Este hecho anecdótico nos recuerda las características de la distribución que lleva su nombre.

Un ejemplo típico en medicina es el número de enfermos que acuden a un servicio de urgencias en una gran ciudad.

La distribución de Poisson como una aproximación de la distribución binomial

La distribución de Poisson es una buena aproximación de la binomial cuando n es grande y p pequeña. En general, cuando n es mayor que 50 y p menor que 0.1 o n x p menor que 5 la distribución de Poisson es una buena aproximación de la distribución binomial.

4. DISTRIBUCIONES DE PROBABILIDAD DE VARIABLES CONTINUAS.

Definición

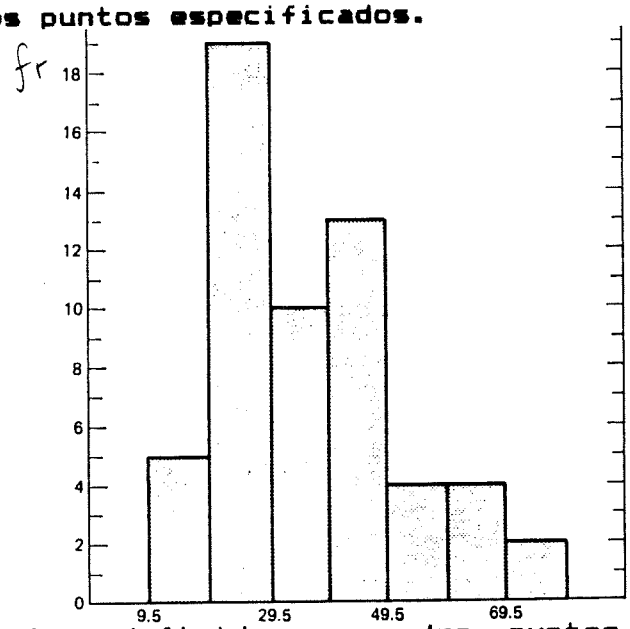
Se define una variable aleatoria continua como aquella que puede tener un número infinito de valores entre dos puntos fijos cualesquiera.

Representaciones gráficas

En la siguiente tabla se muestran 57 valores de una variable aleatoria continua.

1	68	16	65	31	12	46	22
2	63	17	43	32	32	47	43
3	42	18	25	33	49	48	27
4	27	19	74	34	38	49	49
5	30	20	51	35	42	50	28
6	36	21	36	36	27	51	23
7	28	22	42	37	31	52	19
8	32	23	28	38	50	53	46
9	79	24	31	39	38	54	30
10	27	25	28	40	21	55	43
11	22	26	25	41	16	56	49
12	23	27	45	42	24	57	12
13	24	28	12	43	69		
14	25	29	57	44	47		
15	44	30	51	45	23		

El histograma que se muestra a continuación se construyó localizando los puntos especificados sobre una recta que representa la medida de interés y levantando una serie de rectángulos cuyos anchos fueron las distancias entre dos puntos especificados sobre la recta y cuyas alturas representan el número de valores de la variable que caen entre los dos puntos especificados.

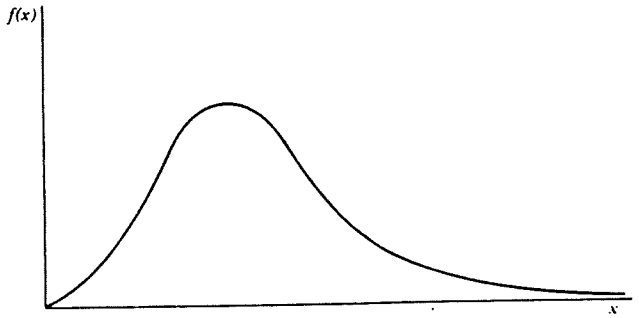


Los intervalos definidos por dos puntos especificados consecutivos reciben el nombre de intervalos de clase. Las subáreas del histograma corresponden a las frecuencias de la ocurrencia de los valores de la variable entre las fronteras de la escala horizontal de estas subáreas. Esto proporciona una manera mediante la cual puede calcularse la frecuencia relativa de la ocurrencia de los valores entre dos puntos

especificados cualesquiera.

Imagínese ahora la situación en donde el número de valores de la variable aleatoria es muy grande y el ancho de los intervalos de clase es muy pequeño.

Si se unieran los puntos medios de las celdas del histograma en la figura anterior para formar un polígono de frecuencias, evidentemente se tendría una figura mucho más suave. En general, a medida que el número de observaciones tiende a ∞ y el ancho de los intervalos de clase tiende hacia 0, el polígono de frecuencias se aproxima a una curva suave como la que se muestra a continuación:



Curvas suaves como ésta son las que se usan para representar gráficamente a las distribuciones de las variables aleatorias continuas.

El área total bajo la curva es igual a 1 y la probabilidad de cualquier valor específico de la variable aleatoria es 0.

Con un histograma, las subáreas de interés pueden encontrarse sumando las áreas representadas por sus celdas. En el caso de una curva suave, debe buscarse un método alternativo para encontrar las subáreas. Este método lo proporciona el cálculo integral. Para encontrar el área bajo una curva suave entre dos puntos cualesquiera, a y b, se integra la función de densidad desde a hasta b. Una función de densidad es una fórmula que se usa para representar la distribución de una variable aleatoria continua. Como se verá posteriormente , para todas las distribuciones continuas que consideremos, existe una manera más fácil de encontrar las áreas bajo las curvas.

5. LA DISTRIBUCION NORMAL.

Introducción

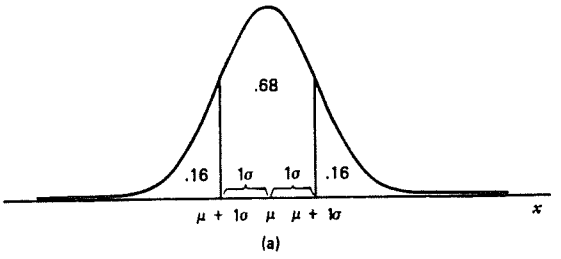
La distribución normal es la distribución más importante de toda la estadística. La fórmula de esta distribución fue publicada por vez primera por Abraham De Moivre (1667-1745) el 12 de noviembre de 1733. Con frecuencia esta distribución se le nombra como gaussiana en reconocimiento a las contribuciones de Carl Friedrich Gauss (1777-1855). La densidad normal está dada por:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2},$$

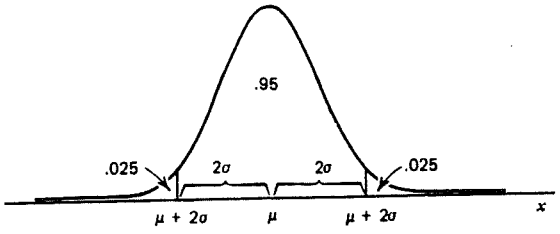
Los dos parámetros de la distribución son la media (μ) y la desviación típica (σ).

Caracteristicas de la distribución

- La media, la mediana y la moda son iguales.
- El área total por debajo de la curva y por encima del eje x es una unidad cuadrada.
- La subdivisión del área bajo la curva normal se muestran en las figuras siguientes:



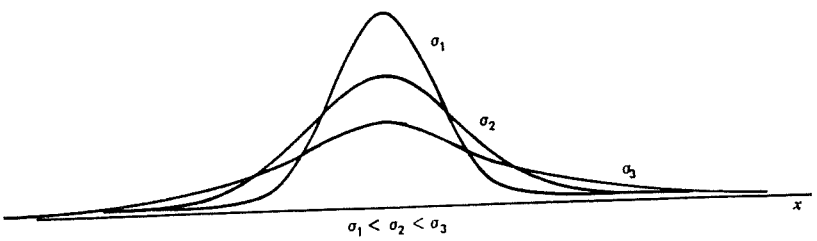
(a)



(b)

- La distribución normal queda completamente determinada por los parámetros μ y σ.

$\mu_1 < \mu_2 < \mu_3$



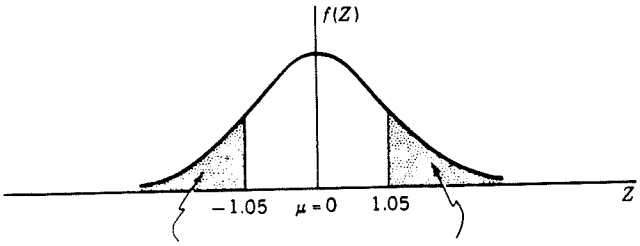
$\sigma_1 < \sigma_2 < \sigma_3$

La distribución normal unitaria

Esta distribución puede obtenerse a partir de la fórmula de la distribución normal general haciendo $\mu=0$ y $\sigma=1$. La variable aleatoria que resulta, $(x-\mu)/\sigma$, generalmente se designa por la letra z. Así la ecuación para la distribución normal unitaria será:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2},$$

Para encontrar la probabilidad de que z tome un valor entre dos puntos cualesquiera del eje z debe encontrarse el área limitada por las perpendiculares levantadas en esos puntos, la curva y el eje horizontal. Afortunadamente, dado que se cuentan con tablas que proporcionan los resultados de todas aquellas integraciones que podrían ser interesantes, no debemos aplicar el cálculo integral. Un ejemplo del mismo se observa en la siguiente figura:



Area = 0.1469 unidades al cuadrado = Area

Area no sombreada = 1 − 2(0.1468) = 0.7062 unidades al cuadrado

Aplicaciones de la distribucion normal

La importancia de la distribución normal se debe principalmente a que hay muchas variables asociadas a fenómenos naturales (distribución de pesos, alturas, coeficientes de inteligencia, puntuaciones en un examen, errores en la medida, etc.) que siguen una distribución aproximadamente normal; también, a que la distribución muestral de muchos estadísticos muestrales, tales como la media, tienen una distribución aproximadamente normal; además, es una buena aproximación de otras distribuciones (así, la distribución de una variable binomial y de Poisson son aproximadamente normales, etc.).

Aunque su importancia en el campo de la estadística es indisputable, se debe tener en cuenta que la distribución normal no es una ley a la que se adhieren todas las características mensurables (medibles) de la naturaleza. Es cierto, sin embargo, que la mayoría de las variables biológicas están distribuidas de forma normal o aproximadamente normal.